Heritability
From Wikipedia, the free encyclopedia

Heritability is the proportion of observed differences on a trait among individuals of a population that are due to genetic differences. Factors including genetics, environment and random chance can all contribute to the variation between individuals in their observable characteristics (in their "phenotypes").[1] Heritability thus analyzes the relative contributions of differences in genetic and non-genetic factors to the total phenotypic variance in a population. For instance, some humans in a population are taller than others; heritability attempts to identify how much genetics play a role in part of the population being taller.

Heritability is measured by estimating the relative contributions of genetic and non-genetic differences to the total phenotypic variation in a population. Heritability is an important concept in quantitative genetics, particularly in selective breeding and behavior genetics (for instance, twin studies), but is less widely used in population genetics.

Heritability measures the fraction of phenotype variability that can be attributed to genetic variation. This is not the same as saying that this fraction of an individual phenotype is caused by genetics. In addition, heritability can change without any genetic change occurring, such as when the environment starts contributing to more variation. A case in point, consider that both genes and environment have the potential to influence intelligence. Heritability could increase if genetic variation increases, causing individuals to show more phenotypic variation, like showing different levels of intelligence. On the other hand, heritability might also increase if the environmental variation decreases, causing individuals to show less phenotypic variation, like showing more similar levels of intelligence. Heritability increases when genetics are contributing more variation or because non-genetic factors are contributing less variation; what matters is the relative contribution. Heritability is specific to a particular population in a particular environment.

The extent of dependence of phenotype on environment can also be a function of the genes involved. Matters of heritability are complicated because genes may canalize a phenotype, making its expression almost inevitable in all occurring environments. Individuals with the same genotype can also exhibit different phenotypes through a mechanism called phenotypic plasticity, which makes heritability difficult to measure in some cases. Recent insights in molecular biology have identified changes in transcriptional activity of individual genes associated with environmental changes. However, there are a large number of genes whose transcription is not affected by the environment.[2]

Contents

- 1 Overview
- 2 Definition
  - 2.1 Example
- 3 Estimating heritability
- 4 Estimation methods
- 5 Regression/correlation methods of estimation
Estimates of heritability use statistical analyses to help to identify the causes of differences between individuals. Since heritability is concerned with variance, it is necessarily an account of the differences between individuals in a population. Heritability can be univariate – examining a single trait – or multivariate – examining the genetic and environmental associations between multiple traits at once. This allows a test of the genetic overlap between different phenotypes: for instance hair color and eye color. Environment and genetics may also interact, and heritability analyses can test for and examine these interactions (GxE models).

A prerequisite for heritability analyses is that there is some population variation to account for. In practice, all traits vary and almost all traits show some heritability.\(^3\) For example, in a population with no diversity in hair color, "heritability" of hair color would be undefined. In populations with varying values of a trait (e.g. see the image on the left), variance could be due to environment (hair dye for instance) or genetic differences, and heritability could vary from 0-100%.
This last point highlights the fact that heritability cannot take into account the effect of factors which are invariant in the population. Factors may be invariant if they are absent and do not exist in the population, such as no one having access to a particular antibiotic, or because they are omni-present, like if everyone is drinking coffee.

**Definition**

Any particular phenotype can be modeled as the sum of genetic and environmental effects:[4]

\[
\text{Phenotype } (P) = \text{Genotype } (G) + \text{Environment } (E).
\]

Likewise the variance in the trait – \( \text{Var}(P) \) – is the sum of effects as follows:

\[
\text{Var}(P) = \text{Var}(G) + \text{Var}(E) + 2 \text{Cov}(G,E).
\]

In a planned experiment \( \text{Cov}(G,E) \) can be controlled and held at 0. In this case, heritability is defined as:

\[
H^2 = \frac{\text{Var}(G)}{\text{Var}(P)}.
\]

\( H^2 \) is the broad-sense heritability. This reflects all the genetic contributions to a population's phenotypic variance including additive, dominant, and epistatic (multi-genic interactions), as well as maternal and paternal effects, where individuals are directly affected by their parents' phenotype, such as with milk production in mammals.

A particularly important component of the genetic variance is the additive variance, \( \text{Var}(A) \), which is the variance due to the average effects (additive effects) of the alleles. Since each parent passes a single allele per locus to each offspring, parent-offspring resemblance depends upon the average effect of single alleles. Additive variance represents, therefore, the genetic component of variance responsible for parent-offspring resemblance. The additive genetic portion of the phenotypic variance is known as Narrow-sense heritability and is defined as

\[
h^2 = \frac{\text{Var}(A)}{\text{Var}(P)}
\]

An upper case \( H^2 \) is used to denote broad sense, and lower case \( h^2 \) for narrow sense.

Additive variance is important for selection. If a selective pressure such as improving livestock is exerted, the response of the trait is directly related to narrow-sense heritability. The mean of the trait will increase in the next generation as a function of how much the mean of the selected parents differs from the mean of the population from which the selected parents were chosen. The observed response to selection leads to an estimate of the narrow-sense heritability (called realized heritability). This is the principle underlying artificial selection or breeding.

**Example**

The simplest genetic model involves a single locus with two alleles (b and B) affecting one quantitative phenotype.

The number of B alleles can vary from 0, 1, or 2. For any genotype, B_B, the expected phenotype can then be written as the sum of the overall mean, a linear effect, and a dominance deviation:
\[ P_{ij} = \mu + \alpha_i + \alpha_j + d_{ij} \]

The additive genetic variance at this locus is the weighted average of the squares of the additive effects:

\[ \text{Var}(A) = f(bb)a_{bb}^2 + f(Bb)a_{Bb}^2 + f(BB)a_{BB}^2, \]

where \( f(bb)a_{bb} + f(Bb)a_{Bb} + f(BB)a_{BB} = 0 \).

There is a similar relationship for variance of dominance deviations:

\[ \text{Var}(D) = f(bb)d_{bb}^2 + f(Bb)d_{Bb}^2 + f(BB)d_{BB}^2, \]

where \( f(bb)d_{bb} + f(Bb)d_{Bb} + f(BB)d_{BB} = 0 \).

The linear regression of phenotype on genotype is shown in Figure 1.

**Estimating heritability**

Since only \( P \) can be observed or measured directly, heritability must be estimated from the similarities observed in subjects varying in their level of genetic or environmental similarity. The statistical analyses required to estimate the genetic and environmental components of variance depend on the sample characteristics. Briefly, better estimates are obtained using data from individuals with widely varying levels of genetic relationship - such as twins, siblings, parents and offspring, rather than from more distantly related (and therefore less similar) subjects. The standard error for heritability estimates is improved with large sample sizes.

In non-human populations it is often possible to collect information in a controlled way. For example, among farm animals it is easy to arrange for a bull to produce offspring from a large number of cows and to control environments. Such experimental control is impossible when gathering human data, relying on naturally occurring relationships and environments.

Studies of human heritability often utilize adoption study designs, often with identical twins who have been separated early in life and raised in different environments (see for example Fig. 2). Such individuals have identical genotypes and can be used to separate the effects of genotype and environment. A limit of this design is the common prenatal environment and the relatively low numbers of twins reared apart. A second and more common design is the twin study in which the similarity of identical and fraternal twins is used to estimate heritability. These studies can be limited by the fact that identical twins are not completely genetically identical, potentially resulting in an underestimation of heritability. Studies of twins also examine differences between twins and non-twin siblings, for instance to examine phenomena such as intrauterine competition (for example, twin-to-twin transfusion syndrome).

Heritability estimates are always relative to the genetic and environmental factors in the population, and are not absolute measurements of the contribution of genetic and environmental factors to a phenotype.

Heritability estimates reflect the amount of variation in genotypic effects compared to variation in environmental effects.

Heritability can be made larger by diversifying the genetic background, e.g., by using only very outbred individuals (which increases \( \text{Var}_G \)) and/or by minimizing environmental effects (decreasing \( \text{Var}_E \)).
converse also holds. Due to such effects, different populations of a species might have different heritabilities for the same trait.

In observational studies, or because of evokative effects (where a genome evokes environments by its effect on them), G and E may covary: gene environment correlation. Depending on the methods used to estimate heritability, correlations between genetic factors and shared or non-shared environments may or may not be confounded with heritability.\[5\]

Heritability estimates are often misinterpreted if it is not understood that they refer to the proportion of variation between individuals in a population that is influenced by genetic factors. Heritability describes the population, not individuals within that population. For example, It is incorrect to say that since the heritability of a personality trait is about .6, that means that 60% of your personality is inherited from your parents and 40% comes from the environment.

A highly heritable trait (such as eye color) assumes environmental inputs which (though they are invariant in most populations) are required for development: for instance temperatures and an atmosphere supporting life, etc. A more useful distinction than "nature vs. nurture" is "obligate vs. facultative"—under typical environmental ranges, what traits are more "obligate" (e.g., the nose—everyone has a nose) or more "facultative" (sensitive to environmental variations, such as specific language learned during infancy).

Another useful distinction is between traits that are likely to be adaptations (such as the nose) vs. those that are byproducts of adaptations (such the white color of bones), or are due to random variation (non-adaptive variation in, say, nose shape or size).

### Estimation methods

There are essentially two schools of thought regarding estimation of heritability.

One school of thought was developed by Sewall Wright at The University of Chicago, and further popularized by C. C. Li (University of Chicago) and J. L. Lush (Iowa State University). It is based on the analysis of correlations and, by extension, regression. Path Analysis was developed by Sewall Wright as a way of estimating heritability.

The second was originally developed by R. A. Fisher and expanded at The University of Edinburgh, Iowa State University, and North Carolina State University, as well as other schools. It is based on the analysis of variance of breeding studies, using the intraclass correlation of relatives. Various methods of estimating components of variance (and, hence, heritability) from ANOVA are used in these analyses.

### Regression/correlation methods of estimation

The first school of estimation uses regression and correlation to estimate heritability.

### Selection experiments

Calculating the strength of selection, $S$ (the difference in mean trait between the population as a whole and the selected parents of the next generation, also called the selection differential\[6\]) and response to selection $R$ (the difference in offspring and whole parental generation mean trait) in an artificial selection experiment will allow calculation of realized heritability as the response to selection relative to the strength of selection, $h^2=R/S$ as in Fig. 3.

### Comparison of close relatives

In the comparison of relatives, we find that in general,
\[ h^2 = \frac{r}{R/S} = \frac{t^2}{r} \]
where \( r \) can be thought of as the coefficient of relatedness, \( b \) is the coefficient of regression and \( t \) the coefficient of correlation.

**Parent-offspring regression**

Heritability may be estimated by comparing parent and offspring traits (as in Fig. 4). The slope of the line (0.57) approximates the heritability of the trait when offspring values are regressed against the average trait in the parents. If only one parent's value is used then heritability is twice the slope. (Note that this is the source of the term "regression," since the offspring values always tend to regress to the mean value for the population, i.e., the slope is always less than one). This regression effect also underlies the DeFries Fulker method for analyzing twins selected for one member being affected.[7]

**Sibling comparison**

A basic approach to heritability can be taken using full-Sib designs: comparing similarity between siblings who share both a biological mother and a father.[8] When there is only additive gene action, this sibling phenotypic correlation is an index of familiarity – the sum of half the additive genetic variance plus full effect of the common environment. It thus places an upper-limit on additive heritability of twice the full-Sib phenotypic correlation. Half-Sib designs compare phenotypic traits of siblings that share one parent with other sibling groups.

**Twin studies**

Heritability for traits in humans is most frequently estimated by comparing resemblances between twins (Fig. 2 & 5). "The advantage of twin studies, is that the total variance can be split up into genetic, shared or common environmental, and unique environmental components, enabling an accurate estimation of heritability".[9] Fraternal or dizygotic (DZ) twins on average share half their genes (assuming there is no assortative mating for the trait), and so identical or monozygotic (MZ) twins on average are twice as genetically similar as DZ twins. A crude estimate of heritability, then, is approximately twice the difference in correlation between MZ and DZ twins, i.e. Falconer's formula \( H^2 = 2(\tau_{MZ} - \tau_{DZ}) \).

The effect of shared environment, \( c^2 \), contributes to similarity between siblings due to the commonality of the environment they are raised in. Shared environment is approximated by the DZ correlation minus half heritability, which is the degree to which DZ twins share the same genes, \( c^2 = DZ-1/2 h^2 \). Unique environmental variance, \( e^2 \), reflects the degree to which identical twins raised together are dissimilar, \( e^2 = 1 - \tau_{MZ} \).
The methodology of the classical twin study has been criticized, but some of these criticisms do not take into account the methodological innovations and refinements described above.

**Extended pedigree design**

While often heritability is analyzed in single generations: comparing MZ twins raised apart, or comparing the similarity of MZ and DZ twins, considerable power can be gained using more complex relationships. By studying a trait in multi-generational families, the multiple recombination of genetic and environmental effects can be decomposed using software such as ASReml and heritability estimated.[10] This design is helpful for untangling confounds such as reverse causality, maternal effects such as the prenatal environment, and confounding of genetic dominance, shared environment, and maternal gene effects[11][12] but is generally less powerful than the twin design for obtaining heritability estimates.

**Analysis of variance methods of estimation**

The second set of methods of estimation of heritability involves ANOVA and estimation of variance components.

**Basic model**

We use the basic discussion of Kempthorne (1957 [1969]). Considering only the most basic of genetic models, we can look at the quantitative contribution of a single locus with genotype $G_i$ as

$$y_i = \mu + g_i + e$$

where

$g_i$ is the effect of genotype $G_i$

and $e$ is the environmental effect.

Consider an experiment with a group of sires and their progeny from random dams. Since the progeny get half of their genes from the father and half from their (random) mother, the progeny equation is

$$z_i = \mu + \frac{1}{2}g_i + e$$

**Intraclass correlations**

Consider the experiment above. We have two groups of progeny we can compare. The first is comparing the various progeny for an individual sire (called *within sire group*). The variance will include terms for genetic variance (since they did not all get the same genotype) and environmental variance. This is thought of as an *error* term.

The second group of progeny are comparisons of means of half sibs with each other (called *among sire group*). In addition to the error term as in the within sire groups, we have an addition term due to the differences among different means of half sibs. The intraclass correlation is

$$\text{corr}(z, z') = \text{corr}(\mu + \frac{1}{2}g + e, \mu + \frac{1}{2}g + e') = \frac{1}{4}V_g,$$

since environmental effects are independent of each other.
The ANOVA

In an experiment with \( n \) sires and \( r \) progeny per sire, we can calculate the following ANOVA, using \( V_g \) as the genetic variance and \( V_e \) as the environmental variance:

**Table 1: ANOVA for Sire experiment**

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>Mean Square</th>
<th>Expected Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among sire groups</td>
<td>( n - 1 )</td>
<td>( S )</td>
<td>( \frac{3}{4}V_g + V_e + r(\frac{1}{4}V_g) )</td>
</tr>
<tr>
<td>Within sire groups</td>
<td>( n(r - 1) )</td>
<td>( W )</td>
<td>( \frac{3}{4}V_g + V_e )</td>
</tr>
</tbody>
</table>

The \( \frac{1}{4}V_g \) term is the intraclass correlation among half sibs. We can easily calculate

\[
H^2 = \frac{V_g}{V_g + V_e} = \frac{4(S - W)}{S + (r - 1)W}.
\]

The Expected Mean Square is calculated from the relationship of the individuals (progeny within a sire are all half-sibs, for example), and an understanding of intraclass correlations.

**Model with additive and dominance terms**

For a model with additive and dominance terms, but not others, the equation for a single locus is

\[
y_{ij} = \mu + \alpha_i + \alpha_j + d_{ij} + e,
\]

where

\( \alpha_i \) is the additive effect of the \( i \)th allele, \( \alpha_j \) is the additive effect of the \( j \)th allele, \( d_{ij} \) is the dominance deviation for the \( ij \)th genotype, and \( e \) is the environment.

Experiments can be run with a similar setup to the one given in Table 1. Using different relationship groups, we can evaluate different intraclass correlations. Using \( V_a \) as the additive genetic variance and \( V_d \) as the dominance deviation variance, intraclass correlations become linear functions of these parameters. In general,

\[
\text{Intraclass correlation} = rV_a + \theta V_d,
\]

where \( r \) and \( \theta \) are found as

\( r = P[ \text{alleles drawn at random from the relationship pair are identical by descent}] \), and

\( \theta = P[ \text{genotypes drawn at random from the relationship pair are identical by descent}] \).

Some common relationships and their coefficients are given in Table 2.
Table 2: Coefficients for calculating variance components

<table>
<thead>
<tr>
<th>Relationship</th>
<th>$r$</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identical Twins</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Parent-Offspring</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
</tr>
<tr>
<td>Half Siblings</td>
<td>$\frac{1}{4}$</td>
<td>0</td>
</tr>
<tr>
<td>Full Siblings</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{4}$</td>
</tr>
<tr>
<td>First Cousins</td>
<td>$\frac{1}{8}$</td>
<td>0</td>
</tr>
<tr>
<td>Double First Cousins</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{16}$</td>
</tr>
</tbody>
</table>

Larger models

When a large, complex pedigree is available for estimating heritability, the most efficient use of the data is in a restricted maximum likelihood (REML) model. The raw data will usually have three or more data points for each individual: a code for the sire, a code for the dam and one or several trait values. Different trait values may be for different traits or for different time points of measurement.

The currently popular methodology relies on high degrees of certainty over the identities of the sire and dam; it is not common to treat the sire identity probabilistically. This is not usually a problem, since the methodology is rarely applied to wild populations (although it has been used for several wild ungulate and bird populations), and sires are invariably known with a very high degree of certainty in breeding programmes. There are also algorithms that account for uncertain paternity.


Response to selection

In selective breeding of plants and animals, the expected response to selection of a trait with known narrow-sense heritability $h^2$ can be estimated using the breeder's equation:[13]

$$R = h^2 S$$

In this equation, the Response to Selection (R) is defined as the realized average difference between the parent generation and the next generation, and the Selection Differential (S) is defined as the average difference between the parent generation and the selected parents.

For example, imagine that a plant breeder is involved in a selective breeding project with the aim of increasing the number of kernels per ear of corn. For the sake of argument, let us assume that the average ear of corn in the parent generation has 100 kernels. Let us also assume that the selected parents produce corn with an average of 120 kernels per ear. If $h^2$ equals 0.5, then the next generation will produce corn with an average of 0.5(120-100) = 10 additional kernels per ear. Therefore, the total number of kernels per ear of
corn will equal, on average, 110.

Note that heritability in the above equation is equal to the ratio \( \frac{\text{Var}(A)}{\text{Var}(P)} \) only if the genotype and the environmental noise follow Gaussian distributions.

**Controversies**

Some authors like Steven Rose\(^{[14]}\) and Jay Joseph\(^{[15]}\) have dismissed heritability estimates as useless. Criticisms have also been raised with regards to the lack of progress in identifying causal genetic variants for many traits (Missing heritability problem). A 2008 paper in *Nature Reviews Genetics* stated however: "Despite continuous misunderstandings and controversies over its use and application, heritability remains key to the response to selection in evolutionary biology and agriculture, and to the prediction of disease risk in medicine. Recent reports of substantial heritability for gene expression and new estimation methods using marker data highlight the relevance of heritability in the genomics era."\(^{[12]}\) According to Eric Turkheimer, new molecular genetic methods of heritability estimation have demonstrated that the underlying assumptions of the classic twin design of human behavioral genetics are sound, "putting an end to an entire misguided school of thought among traditional opponents of classical quantitative (and by association behavioral) genetics."\(^{[16]}\)

**See also**

- Heredity
- Heritability of IQ

**References**

**Notes**

4. Kempthorne 1957


Books


Further reading

External links

- Stanford Encyclopedia of Philosophy entry on Heredity and Heritability (http://plato.stanford.edu/entries/heredity/)
- Quantitative Genetics Resources website, including the two volume book by Lynch and Walsh. Free access (http://nitro.biosci.arizona.edu/zbook/book.html)


Categories: Genetics